DOCUMENT RESUME

ED 178 581 TM 009 747

AUTHOR Bulcock, J. W.; And Others
TITLE The Variance Normalization Method of Ridge Regression Analysis.
PUB DATE 9 Apr 79
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, CA, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation; Foreign Countries; Learning; Multiple Regression Analysis; Sampling; Secondary Education; *Statistical Analysis; Statistical Data
IDENTIFIERS Canada; Multicollinearity; *Ridge Regression

ABSTRACT

The testing of contemporary sociological theory often calls for the application of structural-equation models to data which are inherently collinear. It is shown that simple ridge regression, which is commonly used for controlling the instability of ordinary least squares regression estimates in ill-conditioned data sets, is not a legitimate approach. This is because the minimum mean square error criterion procedure is stochastic; fails to satisfy the boundary condition; and is inadmissible because of a cyclical relationship between the biasing parameter, the inflated residual sum of squares, and the ridge estimator. A new criterion called variance normalization criterion of ridge regression is proposed which is shown to resolve most of the dilemmas associated with the minimum mean square error approach. The several advantages of the new criterion are illustrated through application of a model of school learning in Ontario, Canada high schools. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM 009 747

April 9, 1979

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED178581

THE VARIANCE NORMALIZATION METHOD OF RIDGE REGRESSION ANALYSIS*

J.W. Bulcock
W.F. Lee
Memorial University of Newfoundland
W.S. Luk
Simon Fraser University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J.W. Bulcock

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

*The research was supported by Grant No. S86-786 from the Social Sciences and Humanities Research Council of Canada and Grant No. A4353 from the National Research Council of Canada. We extend thanks to Miss Donna Batten for programming assistance at various stages of the research. An earlier version of this paper was prepared for presentation at the Annual American Educational Research Association Meetings, San Francisco, April, 1979.

TM 009 747

ABSTRACT

The testing of contemporary sociological theory often calls for the application of structural-equation models to data which is inherently collinear. It is shown that simple ridge regression, which is commonly used for controlling the instability of ordinary least squares regression estimates in ill-conditioned data sets, is not a legitimate approach because the minimum mean square error criterion procedure is stochastic, fails to satisfy the boundary condition and is inadmissible because of a cyclical relationship between the biasing parameter "k", the inflated residual sum of squares "ϕ_o", and the ridge estimator "$\hat{\beta}$*". A new criterion called the variance normalization criterion of ridge regression is proposed which is shown to resolve most of the dilemmas associated with the minimum mean square error approach. The several advantages of the new criterion are illustrated through application to a model of school learning.

THE VARIANCE NORMALIZATION METHOD OF RIDGE REGRESSION ANALYSIS

The path of scientific development proceeds from the formulation and estimation of relatively simple models to more complex models. Consider the problems of isomorphism at the interface of auxiliary and statistical levels of modelling. The earliest statistical models - for example, those of Sir Ronald A. Fisher (1925, 1935), and followers - were designed to analyze the data of controlled experiments. These involved very few variables since the potential disturbances of contaminating influences were eliminated through randomization. But not all research problems can be structured to fit the experimental mold. The fact is that the testing of several treatments necessarily restricts the population to that included in the experimental design; and, further, experimental subjects are rarely representative of the population of substantive research interest which means that the preferred realism of the natural setting is seldom present.

In much social science research, then, the chief advantage of experimentation - control through randomization - is usually at the expense of the representativeness of probability samples which permits statistical inferences to the defined population. As Kish (1975) points out, "a research design always involves compromises of the desirable with the possible." But where survey research designs are preferred, efforts to contend with their major limitation - lack of control over potentially disturbing variables - demand isomorphic

statistical models which call for more advanced statistical techniques than those of experimental designs. Thus, survey research methodology is designed to uncover complex networks of relationships which involve the application of statistical controls on potentially confounding variables.

The most efficient methods of removing the effects of a disturbance variable from the mean outcome differences between treatments, or the categories of a predictor variable, have been dependent not only on the development of more advanced mathematical statistics but also on the advent of the computer. One is reminded that these electronic devices are less than forty years old and were conceived by John von Neuman and others less than fifty years ago. In the absence of these two conditions - the development of advanced statistics and the advent of the computer - the processing of survey data and the estimation of more complex models, which call for covariance or regression methods or their nonparametric analogues such as dummy variable regression or multiple classification analysis, would be impossible.

More recent advances in survey research designs by sociologists such as Blalock (1964) and Duncan (1966) have facilitated the development of recursive models of the interdependencies among sets of variables. Parallel advances documented by economists such as Goldberger (1964), Johnston (1963, 1972), and Theil (1971) have centered on the development of nonrecursive simultaneous-equation models for representing

social processes. Both groups use flow graphs or path diagrams for modelling life cycle and other change phenomena as aids both to the conceptualization of model complexity and to the interpretation of their parameter estimates. Of equal importance according to Land and Felson (1978), is that these classes of models are being extensively used in view of the fact that most, though not all, of their known statistical problems have been resolved. For a scholarly review of the structural equation tradition in sociological research see the recent paper by Bielby and Hauser (1978).

Be this as it may, these advances have given rise to a new generation of problems (*sui generis*). For example, they make the specification and estimation of multi-stage interdependent system models and causal chain models (e.g., Tinbergen, 1939) possible. But the better the specification - or, the smaller the residual values of the endogenous variables at each stage - the greater the multicollinearity. Similarly, because social science variables seldom present themselves in the pure or unidimensional form as for example time and mass in physics they are more often than not co-related operationally even though conceptually distinct. Thus, multicollinearity in structural equation models is <u>always</u> a matter of degree not of kind.[1]

In this paper we are concerned with coping more adequately than at present with the multicollinearity problem. First, some of the harmful effects of multicollinearity - a

condition where the smallest eigenvalue of the correlation matrix of an ill-conditioned data set is much smaller than unity (Hoerl and Kennard, 1970a) - are discussed. Second, the properties of simple ridge regression are examined since it is generally believed that this technique based on a minimum total mean square error criterion is the most promising method of coping with the problem. We conclude, however, that this procedure is inadmissible on three grounds, which necessitates the formulation and application of an alternative strategy. The proposed new strategy calls for a different criterion called the variance normalization criterion which resolves all the dilemmas of the former method. Finally, in part four, the advantages of the new ridge regression method are illustrated through application to a model of school learning.

THE MULTICOLLINEARITY PROBLEM[2]

Because models with sets of linearly independent predictor variables are virtually unknown in sociological research, multicollinearity is unavoidable; and so, too, are its six harmful effects. In the first place estimated regression coefficients will have very large sampling variance. The total sampling variance for ordinary least squares (OLS) estimates is given as follows (Hoerl and Kennard, 1970a):

$$\Sigma \mathrm{Var}(\hat{\beta}) = \sigma^2 \mathrm{Tr}(X'X)^{-1} \tag{1}$$

$$= \sigma^2 \Sigma \frac{1}{\lambda_i} > \sigma^2/\lambda_{min} \tag{2}$$

Where Σ represents the summation from i = 1 to i = p, the number of predictors, and λ_{min} is the smallest eigenvalue of the correlation matrix (X'X), which approaches zero for data sets approaching perfect collinearity. Thus, the total variance; and, hence, the sampling variance of individual predictors, can be unacceptably large - even approaching infinity - when the data approaches perfect collinearity.

Second, estimated coefficient vectors, $\hat{\beta}$s may be far too long in ill-conditioned data sets. Geometrically, total variance (Eq. 1) can be regarded as the squared distance from the estimated coefficient vector, $\hat{\beta}$, to the true vector β; and the greater the collinearity the longer the $\hat{\beta}$ to β distance. That is:

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \sigma^2\Sigma\frac{1}{\lambda_i} > \beta'\beta + \sigma^2/\lambda_{min} \qquad (3)$$

And from these two properties it may be concluded that for ill-conditioned data sets there is a high probability that the estimated coefficient vector, $\hat{\beta}$, will be "off" both in magnitude and direction. Thus, the estimated coefficients will be overly sensitive to the sampling error or to the functional misspecification of the model due to the large variance. Obviously, under these conditions the inflated β estimates of predictor equations will not accurately reflect the net relative importance of the variables under consideration. By the same token some non-significant coefficients may in fact be interpreted as important ones because they will be spuriously high. Thus, it is well known by experienced analysts that where multi-

collinearity is a problem an estimated coefficient may be larger than the correlation coefficient between the predictor and dependent variable, and occasionally the OLS regression coefficients may be larger than one.

Third, when the data set is ill-conditioned OLS regression models do not provide accurate prediction; not even when the R^2 is high. Though this follows from the above discussion regarding the instabilities and unreliabilities of regression coefficients due to multicollinearity the conclusion can be formulated explicitly.

Forecasting error variance, σ_f^2, of a model of the form $y = x\beta + \varepsilon$, can be written:

$$\sigma_f^2 = \sigma^2[1 + \underset{\sim}{x}'V\underset{\sim}{x}] \tag{4}$$

where $\underset{\sim}{x}$ is a (p x 1) column vector of a single observation of the predictor variables, X's; V is the variance-covariance matrix of the estimated regression coefficients, which is $(X'X)^{-1}$ for OLS estimates, and σ^2 is the variance of the random error in y. Therefore, for OLS estimates the forecasting error can be expressed as:

$$\begin{aligned}\hat{\sigma}_f^2 &= \hat{\sigma}^2[1 + \underset{\sim}{x}'(X'X)^{-1}\ \underset{\sim}{x}] \\ &= \hat{\sigma}^2[1 + \underset{\sim}{x}^{*1}\Lambda^{-1}\underset{\sim}{x}^*] \\ &= \hat{\sigma}^2[1 + \Sigma\frac{x_i^{*2}}{\lambda_i}] \end{aligned} \tag{5}$$

$$\text{or} \qquad \hat{\sigma}_f^2 = \frac{y'v(1-R^2)}{n-p}[1 + \Sigma\frac{x_i^{*2}}{\lambda_i}] > \frac{y'y(1-R^2)}{n-p}\frac{x_p^{*2}}{\lambda_{min}} \tag{6}$$

where $\underset{\sim}{x}^* = P'\underset{\sim}{x}$ and P is the orthogonal transformation matrix of (X'X), such that $P'(X'X)P = \Lambda$, the eigenvalue matrix of (X'X), and $P'P = PP' = I$. From Eq. (6) it is clear that the forecasting error variance can be very large for ill-conditioned data for which λ_{min} is very small, and that it is more sensitive to the smallest eigenvalue, λ_{min}, than to R^2. This accounts for the fact that high R^2 in OLS regression when the problem has a high degree of multicollinearity does not produce accurate prediction.

Fourth, OLS regression estimates are sensitive to sampling fluctuation when the data set is ill-conditioned. We have already noted (cf. Eq. 2) that for an ill-conditioned data set OLS estimates have large sampling variance due to their inverse relationship with the eigenvalues. Because the sampling variance approaches infinity when the data set approaches perfect collinearity the OLS estimates are sensitive to sampling fluctuation.

Fifth, estimated regression coefficients may be highly dependent; hence, error prone. When a data set has a high degree of multicollinearity, the estimates will have large variance and covariance; consequently they are likely to be highly correlated estimates. If one estimate is erroneous, because of their high dependency, the other will be erroneous. For example, when two predictors are highly positively correlated, the estimated coefficients will generally be highly negatively correlated; which means that if one coefficient is over-estimated, the other will be under-estimated. It is this

high correlation between estimates, plus the large variance of the estimates which accounts for the excessively large sum of squares $\hat{\beta}'\hat{\beta}$ in OLS solutions and why the estimates may even have the wrong sign.

Sixth, OLS estimates are sensitive to model misspecification when the data set is ill-conditioned. Regression coefficients, $\hat{\beta}$'s, are not affected by model misspecification due to the erroneous inclusion or exclusion of variables when the data set is orthogonal. Under this condition the $\hat{\beta}$'s are invariant of the model specification. But when the data set is ill-conditioned, the $\hat{\beta}$'s are highly correlated and, hence, no longer invariant of the model specification, in which case the regression coefficients of the included variables might change drastically as some relevant variables are added to, or dropped from, the model.

Clearly, multicollinearity is an undersirable problem; one to be avoided if at all possible. But as Marquardt and Snee (1975) point out, in the social sciences it is impossible to impose control over all the environmental constraints accounting for intercorrelations between conceptually distinct predictor variables. Since orthogonal explanatory variables are virtually nonexistent; and since, then, OLS procedures will always produce estimates that are inflated, overly sensitive, unreliable, and hard to interpret, a procedure such as simple ridge regression which dampens these harmful effects requires careful examination.

SIMPLE RIDGE REGRESSION

Ridge regression developed by Hoerl and Kennard (1970a, 1970b) is generally accepted as the most promising alternative to OLS regression for coping with collinearity (cf. Dempster et al., 1977). In Hoerl and Kennard's simple ridge regression (SRR) the criterion for measuring the goodness of an estimator $\hat{\beta}^*$ is the total mean square error (MSE)[3] function defined by

$$MSE = E[(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)]. \tag{7}$$

MSE is the sum of the mean square error of the individual regression coefficients, and it can be proved (e.g., Pindyck and Rubinfeld, 1976, p. 22) that it can be decomposed into the sum of the total variance and the total squared bias of each regression coefficient, that is:

$$MSE = \Sigma Var(\hat{\beta}_i^*) + Bias^2\ (\hat{\beta}^*). \tag{8}$$

Since, geometrically, MSE represents the expected squared distance between $\hat{\beta}^*$ and the true vector β, a "good" estimator would be the one that minimizes this distance - that is, minimizes the MSE. And since the OLS estimator, $\hat{\beta}$, gives the minimum residual sum of squares; $\hat{\beta}^*$, the estimator based on the minimum MSE criterion, will give an inflated residual sum of squares, ϕ, which is written as:

$$\begin{aligned} \phi = \varepsilon'\varepsilon &= (y - X\hat{\beta}^*)'\,(y - X\hat{\beta}^*) \\ &= (y - X\hat{\beta})'\,(y - X\hat{\beta}) + (\hat{\beta}^* - \hat{\beta})'\,X'X(\hat{\beta}^* - \hat{\beta}) \\ &= \phi_{min} + \phi_o(\hat{\beta}^*) \end{aligned} \tag{9}$$

where $\phi_o(\hat{\beta}^*) = (\hat{\beta}^* - \hat{\beta})'\,X'X(\hat{\beta}^* - \hat{\beta})$ (10)

is the inflation in the residual sum of squares. Geometrically it gives the surface of a family of hyperellipsoids centered at $\hat{\beta}$, which is the OLS estimate of β.

$\hat{\beta}^*$, then, is the ridge estimator which minimizes the squared length of the coefficient vector which is subjected to constant inflation in the residual sum of squares, $\phi_o(\hat{\beta}^*)$, determined by the minimum MSE. That is, $\hat{\beta}^*$, is a solution to minimizing $\hat{\beta}^{*\prime}\hat{\beta}^*$ subjected to $(\hat{\beta}^* - \hat{\beta})' X'X (\hat{\beta}^* - \hat{\beta}) = \phi_o$ (a constant determined by the minimum MSE).[4]

The solution of this problem is obtained by minimizing the Lagrangian function

$$f(\hat{\beta}^*) = \hat{\beta}^{*\prime}\hat{\beta}^* + \frac{1}{k}[(\hat{\beta}^* - \hat{\beta})' X'X(\hat{\beta}^* - \hat{\beta}) - \phi_o] \qquad (11)$$

where $\frac{1}{k}$ is the Lagrangian multiplier. At minimum $f(\hat{\beta}^*)$, we have

$$\frac{\partial}{\partial \hat{\beta}^*} f(\hat{\beta}^*) = 2\hat{\beta}^* + \frac{1}{k} 2X'X(\hat{\beta}^* - \hat{\beta}) = 0$$

$$(X'X + kI)\hat{\beta}^* = X'X\hat{\beta} = X'y$$

therefore,

$$\hat{\beta}^* = (X'X + kI)^{-1} X'y \qquad (12)$$

where k is a parameter determined by Eg. (10), which in turn is determined by minimum MSE (See Hoerl and Kennard, 1970a).

It is clear that the minimum square length of the coefficient vector, or the minimum sum of squared regression coefficients determine the form of the ridge estimator and the minimum MSE determines the value of the parameter k. Note that the form of the ridge estimator $\hat{\beta}^*$ in Eq. (12) is the same as that derived by Lindly and Smith (1972) using Bayesian methods under the assumption of an exchangeable prior distribution for β.

Most properties of ridge regression have been discussed by researchers such as Hoerl and Kennard (1970a) and Marquardt (1970). Recently, Lee (1979) summarized sixteen of them. Here we are only interested in the three which are problematic. First, the ridge estimator $\hat{\beta}^*$ has a total mean square error of:

$$MSE = \sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \Sigma \frac{\alpha_i^2}{(\lambda_i + k)^2} \qquad (13)$$

where $\alpha = P'\beta$ is the true regression coefficient vector in canonical form $y = X^*\alpha + \varepsilon$, and where $X^* = XP$.[5] The first component of MSE is the total variance of estimation which describes the random portion of the error, while the second component is the square bias. It is clear from this relation (Eq. 13) that the total MSE of ridge regression depends on the true unknown parameters σ^2 and α; thus, it cannot be obtained.

Second, for any problem there exists a positive "optimal k" (called k_o) which gives a minimum MSE.[6] It is true even for an orthogonal system for which the degree of multicollinearity is zero. Third, for any problem the optimal k depends on the true regression coefficient vector, β, and the variance of the residual of the linear model; i.e., σ^2.[7] But the multicollinearity problem arises from the interdependency among the explanatory variables and not from the dependency between the dependent variable y, and the explanatory variables X's. But k depends on β and σ^2 which are dependent on the y variable; which means that k is stochastic when in

fact it should be a nonstochastic parameter.

The four dilemmas of ridge regression based on the minimum MSE criterion stem from these three properties; namely, (i) the theoretical value of optimal k is stochastic while it should be nonstochastic if the task is to reduce the harmful effects of multicollinearity, (ii) the true optimal k depends on population parameters and renders the problem unsolvable, (iii) even for an orthogonal system there is an optimum k at which MSE is minimum, and (iv) the bias of the ridge estimates can never be obtained. Thus, the optimum k cannot be calculated and consequently the performance of ridge regression cannot be evaluated. So far more than fifteen methods have been described; for example, Hoerl and Kennard (1970a), Hoerl, Kennard and Baldwin (1975), Vinod (1976), Obenchain (1975), Hocking et al. (1976), McDonald and Galarneaus (1975), Kasarda and Shih (1977), Hemmerle (1975), Hemmerle and Brantle (1978), Guilkey and Murphy (1975), Lawless and Wang (1976), Allen (1974). Unfortunately despite their individual advantages none can give a better k or even a smaller MSE to that of OLS regression - a difficulty that has marred the superiority of ridge regression over the OLS regression procedure.[8]

THE VARIANCE NORMALIZATION METHOD OF RIDGE REGRESSION ANALYSIS

The total variance of the estimated regression coefficients based on the OLS procedure can be expressed as: (see Hoerl and Kennard, 1970a).

$$\Sigma \mathrm{Var}(\hat{\beta}^*) = \hat{\sigma}^2 \mathrm{tr}(X'X)^{-1}$$

$$= \frac{\varepsilon'\varepsilon}{n-p} \mathrm{tr}(X'X)^{-1} \qquad (18)$$

where ε is the random error vector of y in the standardized regression model, $y = X\beta + \varepsilon$, n is the sample size, p is the number of explanatory variables, and $\mathrm{tr}(X'X)^{-1}$ is the trace of the inverse of the correlation matrix (X'X). It is obvious from this expression that the total variance depends on the following three independent factors.

(1) The random error. This is the purely random portion of the dependent variable y. It generally consists of two parts; that due to measurement error ε_m, and stochastic error ε_s due to the incompleteness of the designed model, and some unknown inherent irreproducible fluctuation (cf. Wannacott, 1969, p. 17). Since these two errors are assumed to be uncorrelated and to have normal distribution with zero mean, the sum of squares of random error $\varepsilon'\varepsilon$ can be expressed as

$$\varepsilon'\varepsilon = \varepsilon_m'\varepsilon_m + \varepsilon_s'\varepsilon_s.$$

This random error can be reduced but not eliminated by improving the precision of measurement and by improving the designed model.

(2) The sample size, n. More specifically, this is the degree of freedom of the sum of the squares error (n-p). Total variance can be reduced by increasing the sample size.

(3) The degree of multicollinearity. The effect of multicollinearity enters through $\mathrm{tr}(X'X)^{-1}$, which is the trace of $(X'X)^{-1}$. Specifically, it is the sum of the variance inflation

factor (VIF) of each variable due to its interrelationship with the rest of the explanatory variables. In any statistical estimation, it is desirable to have the variance as low as possible.

From this discussion it is clear that for classical multiple regression the total variance of the estimated coefficient may be reduced by improving the precision of measurement, by improving model specification, by increasing the sample size, and most important by reducing the inflation of variance due to multicollinearity. The ridge regression procedure developed by Hoerl and Kennard (1970a) was originally designed to perform this last task - reduce the inflation of variance attributable to multicollinearity. But use of the MSE criterion for determining k is effectively a multifunctional procedure because in suppressing the VIF to less than unity it suppresses at the same time the variance attributable to other causes of variance - measurement error, model misspecification, and small sample size. It is this omnibus nature of the method which has forced ridge regression to be a stochastic procedure, and why even for an orthogonal system there is an optimum k which generates minimum MSE.

From the above argument, it follows that, if we want to limit the simple ridge procedure to the performance of the simple task of reducing the inflation of variance due to multicollinearity, the lowest permissible total variance is $p\hat{\sigma}^2$, which is the value for the best condition; namely, that

present in an orthogonal system. That is:

$$\Sigma \mathrm{Var}(\hat{\beta}^*) = p\hat{\sigma}^2$$

$$\hat{\sigma}^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = p\hat{\sigma}^2$$

and therefore,

$$\frac{1}{p} \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} = 1 \qquad (19)$$

In words we could say that in order for SRR to perform the single task of reducing variance inflation due to multicollinearity, and only multicollinearity, one should normalize the average VIF such that it is equal to one. The value of k (called k_N) which satisfies this condition is the k which satisfies Eq. (19).

If after using this procedure, the resultant variance is still too large, and if it is desirable to further reduce the variance, it should be accomplished through improvement in measurement error, model specification and sample size; but not by further suppression of the variance inflation factor; and especially if the intention is the interpretation of the phenomenon. In the tail ($k > k_N$) portion of the trace, the "trade-off" is not efficient and would introduce, therefore, too large a bias.

It has been shown elsewhere (Lee, 1979; Lee and Luk, 1979) that the normalization procedure is more consistent in terms of optimum performance than six other ridge regression methods (viz., those of Kasarda and Shih, 1977; Hocking, Speed and Lynn, 1976; Lawless and Wang, 1976; Hoerl, Kennard, and

Baldwin, 1975; and Hoerl and Kennard's iterative, 1976). To be more specific, it was shown that across five problem data sets selected from the ridge regression literature[9] the normalization method: (i) consistently and optimally reduced the maximum VIF to a value close to unity; (ii) consistently generated the maximum reduction of the relative degree of multicollinearity of the problem without introducing too large a bias; (iii) consistently produced a large amount of reduction in the total variance of the estimates; (iv) consistently produced more stable estimates (on the basis of a sensitivity ratio) than OLS regression; (v) generally produced low to acceptably-low inflation in the residual sum of squares. By the same token the other methods were generally not consistent and failed to maximize the reduction of the harmful effects of the multicollinearity problem.

A MODEL OF SCHOOL LEARNING

We agreed in the introduction that scientific advance is from the formulation and estimation of simple to more complex models; that the development of structural equation models including the nonrecursive simultaneous equation variety has made more complex model building possible but that such models have to cope with system level problems which are sui generis. One of the more serious and ubiquitous problems is that of multicollinearity. In this section of the paper the multicollinearity problem is illustrated by reference to a model of school learning (Bulcock et al., 1979).

School learning models at the secondary school level are designed to explain the mutually reinforcing cognitive and affective behaviours of students. These learning outcomes are determined by entry behaviours. The simple argument is that the student's achievement now (the time of testing) is a function of his prior knowledge and attitudes. The effects of entry behaviours are both direct and indirect - the indirect effects are mediated by exposure to what is learned (the opportunity to learn) and how it is taught (the quality of instruction). It is also posited that antecedent considerations or background factors affect the entry behaviours; which with a few exceptions mediate the impact of the antecedent factors on the learning outcomes.

Two additional notions are central to the model specification. First, the thesis is held that the affective characteristics of students - their attitudes and motivations vis-a-vis schooling - affect their success in school. Second, these cognitive behaviours and affective characteristics constitute mutually reinforcing learning systems. Thus, the design of learning models has to incorporate both the affective traits of learners and their feedback properties over and above earlier self-evident notions that achievement stimulates achievement.[10]

The structure of the model of school learning to be estimated is presented as Figure 1. The three antecedent variables were socioeconomic status, sex, and mental ability.

The entry behaviours were the students' levels of literacy and numeracy and their motivations and expectations measured in grade nine. The learning outcomes of interest were their grade ten cognitive performances in English language, and mathematics and their grade ten affective characteristics; namely, motivations and expectations. The model was mis-specified in that the mediating variables - opportunity to learn and instructional processes - were not available in the selected data set. It was hypothesized that the cognitive and affective learning outcomes would constitute a mutually reinforcing system as depicted in Figure 1. With the exception of sex, grade ten English and grade ten mathematics, all variables were latent constructs with alpha reliabilities greater than 0.80.

Figure 1 About Here

The data were obtained from the Carnegie Human Resources Data Bank - a five year panel study of all students (90,000 approximately) who began grade nine in Ontario High Schools in 1959-60. Annual follow-up data was gathered until the students entered University in 1964. The bank holds 90,000 records on Canadian students with upward of 400 pieces of information on the student, his family, his teachers, his community and educational career through five years of high school. The 886 students in the present study were selected as follows. A random sample of 9000 students was selected from the total.

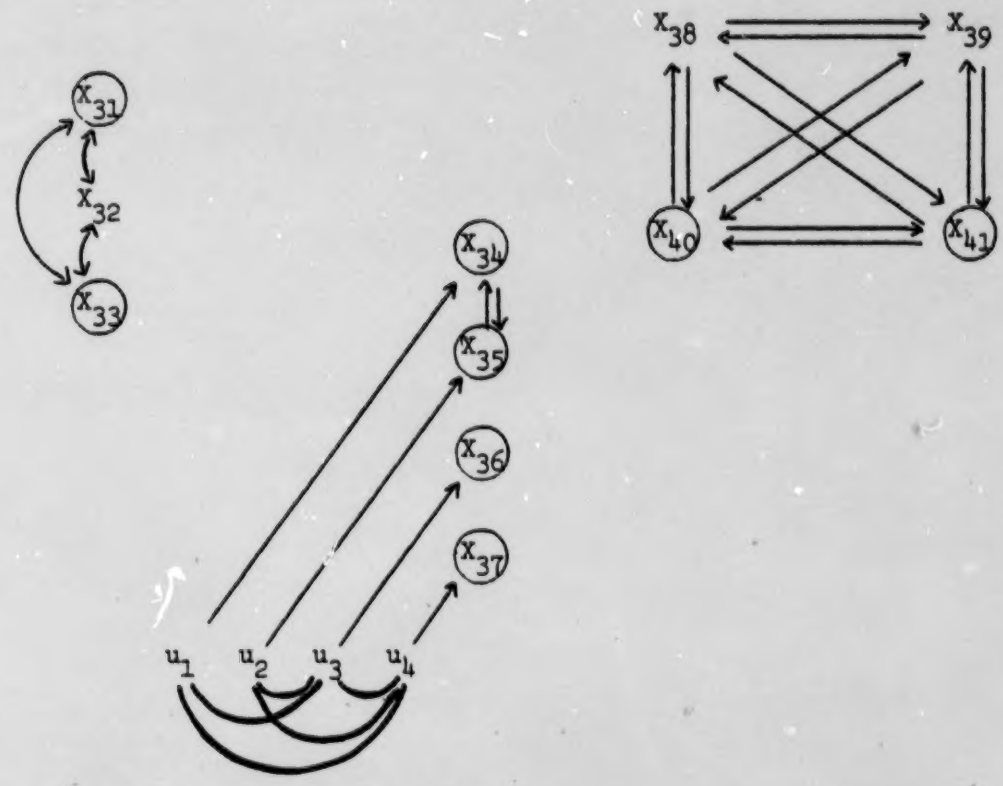

Figure 1. Structure of the Model of School Learning. Recursive path arrows are omitted in the interests of clarity. Variables are assumed to be caused by variables to the left of them. The variable identifications are: X_{31} SES, X_{32} Sex, X_{33} IQ, X_{34} Literacy, X_{35} Numeracy, X_{36} Motivation 1, X_{37} Expectations 1, X_{38} English, X_{39} Mathematics, X_{40} Motivation 2, X_{41} Expectations 2. Latent variables are circled.

After screening out those students who did not speak English at home we were left with a list-wise N of 886. The majority of coefficients in the list-wise (886) and pair-wise (7,809) correlation matrices were the same to two decimal places. There were exactly 443 males and 443 females in the sample analyzed.

We present the comparisons between OLS estimates and SSR estimates with the k_N criterion for a recursive model in Table 1. In the entry behaviours segment of the model (the first four columns) the degree of multicollinearity was modest

---
Table 1 About Here
---

as may be gauged from the Table 2 indices. Thus, the SSR(k_N) estimates were about the same in magnitude, direction and significance as the OLS regression estimates. This is as it should be since the boundary condition - namely, that SSR(k_N) estimates for orthogonal data be the same as OLS estimates - is satisfied by the variance normalization method of ridge regression analysis.

The Table 2 indices require clarification. Though several different indicators of the severity of multicollinearity have been used by different researchers (cf. Price, 1977; Farrar and Glauber, 1976; White, 1978) Marquardt and Snee (1975) suggest that the maximum variance inflation factor, V_{max}, is the best single measure. It is 1 for no collinearity and infinity for perfect multicollinearity. In

Table 1

Path Coefficients and T-values for the OLS and SRR(k_N) Estimates of the Recursive Model of School Learning*

Independent Variables		Dependent Variables: X_{34} LIT(9) B	T	X_{35} NUM(9) B	T	X_{36} MOT1 B	T	X_{37} EXP1 B	T	X_{38} ENG B	T	X_{39} MATH B	T	X_{40} MOT2 B	T	X_{41} EXP2 B	T
X_{31}	SES	.066	2.91	-.012	-.52	.086	2.75	.173	5.58	.023	1.21	.016	.72	-.026	-1.03	.138	6.04
		(.054)	2.76	(-.016)	-.68	(.086)	2.70	(.175)	5.53	(.015)	.68	(.011)	.44	(-.034)	-1.18	(.146)	5.48
X_{32}	SEX	.139	6.32	-.076	-3.41	.183	6.04	-.101	-3.35	.165	8.80	-.071	-3.14	.070	2.81	-.010	-.46
		(.143)	6.36	(-.077)	-3.38	(.187)	6.05	(-.103)	-3.33	(.179)	8.05	(-.079)	-2.95	(.066)	2.23	(-.003)	-.11
X_{33}	IQ	.706	31.27	.715	31.20	.311	9.99	.289	9.33	.226	9.77	.273	9.77	-.004	-.14	.121	4.30
		(.721)	31.22	(.730)	31.16	(.318)	9.97	(.294)	9.29	(.237)	6.50	(.326)	7.43	(-.046)	-.95	(.136)	3.07
X_{34}	LIT(9)									.367	16.40	.103	3.82	.132	4.42	.074	2.72
										(.427)	13.36	(.075)	1.94	(.157)	3.69	(.056)	1.44
X_{35}	NUM(9)									.093	4.20	.224	8.34	.101	3.42	.076	2.80
										(.067)	2.11	(.230)	6.01	(.113)	2.68	(.069)	1.78
X_{36}	MOT1									.117	5.96	.154	6.49	.378	14.44	.095	3.96
										(.117)	4.93	(.172)	6.02	(.433)	13.76	(.095)	3.31
X_{37}	EXP1									.076	3.87	.011	.45	.037	1.42	.372	15.64
										(.073)	3.1[illegible]	(-.003)	-.10	(.030)	.980	(.426)	15.02
R^2		.5320		.5194		.1426		.1532		.5480		.3983		.2722		.3782	
		(.5531)		(.5403)		(.1482)		(.1582)		(.6182)		(.4458)		(.3247)		(.4343)	
Residual		.684		.693		.926		.920		.672		.776		.853		.788	
		(.668)		(.678)		(.923)		(.917)		(.618)		(.744)		(.822)		(.752)	
K_N		.020		.020		.020		.020		.138		.138		.138		.138	

*The OLS regression estimates are in parentheses below the SRR(K_N) estimates in each instance.

order to obtain a measure which gives 0 for no collinearity and 1 for perfect collinearity, we use a D_{max}-measure which is a transformation of V_{max} given by

$$D_{max} = \frac{2}{\pi} \tan^{-1}(V_{max} - 1) \qquad (20)$$

where the arc tangent is measured in radians. D_{max} is the degree of multicollinearity of that variable which is affected the most; hence, its use to represent the severity of multicollinearity in the entire data set. An absolute measure, D_c, of multicollinearity can be defined in a similar way by using the VIF_{max} of the principal component, $1/\lambda_{min}$; that is,

$$D_c = \frac{2}{\pi} \tan^{-1}(\frac{1}{\lambda min} - 1) \qquad (21)$$

This measure represents the absolute degree of multicollinearity of the data set and does not change with the data handling technique.

The D_{max}-measure for the stage two, learning outcomes, equations (the last four columns of Table 1) is far too high to be acceptable. The evidence from Table 1 shows that in general the OLS estimates are too large; and in three instances relationships would have been misinterpreted as insignificant when in fact they were statistically significant (cf. $P_{39,34}$, $P_{41,35}$, and $P_{41,34}$).

In order to reduce the variance inflation due to multicollinearity it is desirable to normalize the average VIF such that it is equal to one. The value of k_N which satisfies the condition at the learning outcomes stage of the recursive model is 0.138. After application of the ridge

constant (k_N = .138) to the stage two matrix the multicollinearity indices may be recalculated as shown in the lower half of Table 2. There is a noticeable reduction in multicollinearity in the SRR(k_N) model compared to that in the OLS regression model.

Table 2 About Here

School learning models call for reciprocal effect relationships at both the entry behaviour and learning outcomes stages (see Figure 1). Two stage least squares (2SLS) results for Figure 1 relationships are compared to two stage ridge regression (2SRR) results in Table 3. The respecification of the recursive model through incorporation of reciprocal paths aggravates the multicollinearity problem.

At the entry behaviour stage of the recursive model there was virtual absence of multicollinearity - the k_N value was a modest 0.02. At the same stage of the reciprocal effects model where it is argued that performance in literacy will stimulate achievement in numeracy and vice versa, the k_N values are .115 and .118 respectively. Further, the 2SLS findings are absurd in terms of the theory.

At the learning outcomes level in the reciprocal effects model the multicollinearity problem is more acute than at the entry behaviour stage. In fact the Table 3 results illustrate every problem of multicollinearity referred to in our above discussion of the problem. The 2SLS T-values indicate that standard errors are unacceptably high; several negative path

Table 2

Indices of the Degree of Multicollinearity in Stage 1 and Stage 2 OLS Equations and Stage 1 and Stage 2 SRR(k_N) Equations

OLS Equations	V_{max}	D_{max}	D_c
Stage 1 - Entry Behaviours	1.066	.042	.199
Stage 2 - Learning Outcomes	3.087	.716	.830
SRR(k_N) Equations	**V_{max}**	**D_{max}**	**D_c**
Stage 1 - Entry Behaviours	1.012	.008	.199
Stage 2 - Learning Outcomes	1.250	.156	.830

estimates are of dubious value; and the magnitude of several estimates is suspiciously high. Consider the three following problems illustrated by Table 3 results.

(i) Compare 2SLS and 2SRR estimates for $p_{38,33}$, $p_{38,40}$, $p_{39,36}$, $p_{40,38}$, and $p_{41,39}$. The 2SLS $\hat{\beta}$'s are too long (Eq. 3) with the result that they are "off" both in magnitude and direction; hence, are of the wrong sign in each instance.

(ii) The following 2SLS estimates - $p_{38,32}$, $p_{38,39}$, $p_{39,40}$, and $p_{40,41}$ - would have been interpreted as important. But after correction for collinearity it is easily seen that they were in fact spuriously high.

(iii) By the same token the following 2SLS estimates - $p_{38,40}$, $p_{40,34}$, $p_{40,38}$, and $p_{41,39}$ - were spuriously low, and would have been interpreted as non-significant when in fact they were not, as the 2SRR estimates and T-values indicate.

Table 3 About Here

CONCLUSION

We have presented analyses for a single example in this paper. Actually many more problems have been analyzed - see the references to some of our earlier work. The results of the present analysis and those in all our other analyses

Table 3

Path Coefficients and T-values for 2SRR and 2SLS
Regression Analyses of the Model of School Learning *

Independent Variables	Dependent Variables							
	X_{34} T LIT(9)	X_{35} T NUM(9)	X_{36} T MOT1	X_{37} T EXP1	X_{38} T ENG(10)	X_{39} T MATH(10)	X_{40} T MOT2	X_{41} T EXP2
X_{31} SES	.072 3.44 (.058) 2.32	-.016 -.75 (-.013) -.51	.086 2.75 (.086) 2.70	.173 5.58 .175 5.53	.002 .114 (-.128) -1.80	.004 .20 (.070) 1.34	-.053 -2.22 (-.112) -2.72	.147 6.71 (.194) 4.65
X_{32} SEX	.145 7.14 (.122) 4.50	-.100 -4.86 (-.062) -2.33	.183 6.04 (.187) 6.05	-.101 -3.35 -.103 -3.33	.165 8.82 (.308) 4.32	-.107 -4.80 (-.160)-3.68	.066 2.79 (.137) 2.67	-.042 -1.91 (-.166)-2.27
X_{33} IQ	.422 21.43 (.902) 7.23	.484 19.96 (.787) 11.04	.311 9.99 (.318) 9.97	.289 9.33 .294 9.29	.142 6.62 (-.210) -1.22	.229 8.49 (.338) 5.15	-.076 -2.69 (-.269) -2.82	.060 2.37 (.287) 1.96
X_{34} LIT	—	.254 8.17 (-.080) -.86			.307 13.56 (.438) 5.61	.037 1.41 (-.091)-1.12	.071 2.51 (.138) 1.87	-.011 -.42 (-.193)-2.07
X_{35} NUM	.343 12.98 (-.246) -1.47	—			.019 .83 (-.136) -1.37	.191 7.27 (.169) 3.31	.040 1.43 (-.066) -.87	.025 .97 (.072) .75
X_{36} MOT1					.009 .47 (.221) 1.44	.054 2.40 (-.130)-1.15	.317 12.56 (.299) 5.59	-.020 -.84 (-.323)-2.77
X_{37} EXP1					-.003 -.15 (-.230) -1.87	-.021 -.94 (.093) .99	-.039 -1.60 (-.179) -2.31	.346 15.07 (.386) 8.75
X_{38} ENG(10)					—	.115 3.73 (.175) 1.30	.082 2.57 (-.124) -.87	.127 4.15 (.269) 1.63
X_{39} MATH(10)					.206 8.75 (.994) 2.38	—	.172 5.69 (.629) 2.32	.080 3.05 (-.580)-1.44
X_{40} MOT2					.161 5.46 (-.776) -1.70	.191 5.29 (.690) 2.49	—	.225 6.08 (1.106) 3.51
X_{41} EXP2					.181 6.25 (.733) 2.58	.063 1.92 (-.277)-1.25	.160 4.30 (.483) 2.84	—
R^2	.5473 (.4925)	.5406 (.5225)	.1426 (.1482)	.1532 (.1582)	.6107 (-.3395)	.4520 (.2154)	.3681 (.1762)	.4662 (-.2338)
Residual	.673 (.712)	.678 (.691)	.926 (.923)	.920 (.917)	.624 (1.157)	.740 (.886)	.795 (.908)	.731 (1.111)
SRR(K_N)	.115	.118	.020	.020	.154	.160	.163	.160

*2SLS regression estimates are in parentheses below the 2SRR estimates in each case. Correlations between antecedent variables are: $r_{31,32}$ = -.042, $r_{31,33}$ = .243, $r_{32,33}$ = -.062. The correlations between entry behaviours residuals (u_1, u_2, u_3, u_4) are: r_{23} = .246, r_{34} = .193, r_{24} = .073, r_{13} = .166, and r_{14} = .173. Variable identifications are: X_{31} SES, X_{32} Sex, X_{33} IQ, X_{34} Literacy, X_{35} Numeracy, X_{36} Motivation 1, X_{37} Expectations 1, X_{38} English, X_{39} Mathematics, X_{40} Motivation 2, X_{41} Expectations 2.

show that the variance normalization method of ridge regression analysis consistently produces the most satisfactory results judged by the maximum reduction in the effects of multicollinearity. Further, the variance normalization criterion avoids all the dilemmas attributable to the application of the minimum MSE criterion of simple ridge regression. With the k_N criterion OLS regression is the boundary condition in recursive models; thus, any problem with any degree of multicollinearity may be analyzed by using ridge regression. Though the results may be the same as OLS results when the degree of multicollinearity is low as was shown in the first stage of the recursive model (Table 1), the normalization method still minimizes the probability of getting "bad" results. The same holds with two stage ridge regression - perhaps more so in view of the fact that reciprocal models aggravate the multicollinearity problem.

Footnotes

1. These advances in structural equation model building were accompanied by parallel software developments. Two well-known alternative, though complementary, general computer packages have been developed by the Swedish econometrician Herman Wold and his former student Karl G. Jöreskog; known as Partial Least Squares (PLS) and Maximum Likelihood Linear Structural Relations (ML-LISREL) respectively. Both sets of statistical procedures allow the user to conduct recursive and nonrecursive path analysis with latent variables in models involving causal feedback and interdependencies. See Wold, 1975, 1977a, 1977b, 1977c, 1978; Hui, 1978; Noonan and Wold, 1977, 1978 for discussions of PLS estimating procedures. See Joreskog 1970, 1973, 1976, 1978 for references to the ML-LISREL mode of estimation. But insofar as the inner relationships are concerned neither procedure is invariant with respect to collinearities, or of the model specification (see property number six of ill-conditioned data sets above). Differences between the two statistical packages include the following. PLS modelling is at an earlier stage of development. It tends to be more open-ended; open, that is, to extension, elaboration and consolidation. Woldian statistical methods are versatile with regard to macro-level, multidisciplinary models which are at the formulation-testing phase of development; and especially when little prior knowledge is available about the residuals. While these purposes can be served with ML-LISREL as a facet of the procedure for the structural analysis of covariance and correlation matrices, the LISREL approach is probably most versatile where the microanalysis of experimental and quasi-experimental data is of interest; and where, usually, prerequisite information about the residuals and their distribution is available. Neither PLS nor LISREL copes satisfactorily with ill-conditioned data sets though little work on this problem is available. Some inferences may be drawn from the findings of Monte Carlo simulation experiments such as the following:

(i) that multicollinearity produces greater bias in two-stage least squares (2SLS) estimates than in OLS estimates (Johnston, 1972, p. 409);

(ii) that in the presence of multicollinearity 2SLS performs marginally better than full information maximum likelihood (FIML) and several other estimating procedures (Summers, 1965; as cited in Johnston 1972, p. 419); but

(iii) that when the predictors are relatively orthogonal FIML is a better estimating procedure than 2SLS;

(iv) that most complex simultaneous-equation procedures are more sensitive to specification errors than 2SLS (Blacock, in Namboodiri et al., 1975, p. 517); and

(v) that 2SLS is at least as good as three-stage least squares, but difficulties occur when exogenous variables are highly intercorrelated (King, 1974).

Some of our recent work demonstrates the limitations of these current structural equation methods - especially OLS regression and 2SLS regression - when multicollinearity is an inherent property of the data set (cf. Beebe and Bulcock, 1979; Bulcock and Lee, 1979; Bulcock, Lee, and Luk, 1979; Lee and Luk, 1979). We find that some of the most interesting school learning models, psycholinguistic models of reading comprehension, and socio-economic career models cannot be estimated adequately because longitudinal designs calling for feedback relationships are inherently collinear.

2. The problems of multicollinearity have been discussed elsewhere; see, for example, Althauser (1971), Blalock (1963, 1964), Christ (1966), Gordon (1968), Hoerl and Kennard (1970a), Johnston (1972), Kasarda and Shih (1977), Kmenta (1971), Marquardt (1970), Marquardt and Snee (1975), and Rockwell (1975). None brings out the overall magnitude and extent of the problem to our satisfaction.

3. Of the two MSE criteria used in ridge regression, weighted and unweighted (cf. Hocking et al., 1976), only the unweighted one is considered in this paper.

4. It is pointed out in the next section of this paper that in the variance normalization method of ridge regression, ϕ_o, is determined indirectly by normalizing the average variance inflation factor (VIF).

5. Proof: the total mean square error

$$\begin{aligned} MSE &= E(\hat{\beta}^* - \beta)^2 = E\{[\hat{\beta}^* - E(\hat{\beta}^*)] + [E(\hat{\beta}^*) - \beta]\}^2 \\ &= E[\hat{\beta}^* - E(\hat{\beta}^*)]^2 + [E(\hat{\beta}^*) - \beta]^2 \\ &= \text{tr cov}(\hat{\beta}^*) + \text{Bias}^2(\hat{\beta}^*) \\ &= \sigma^2 \text{ tr } P \text{ Diag}[\lambda_i/(\lambda_i + k)^2]P' + k^2 \alpha' \text{ Diag}[1/(\lambda_i + k)^2]\alpha \\ &= \sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \Sigma \frac{\alpha_i^2}{(\lambda_i + k)^2} \end{aligned}$$

6. Proof (following Kasarda and Shi, 1977): Since

$$MSE = \Sigma \text{Var}(\hat{\beta}_i^*) + \text{Bias}^2(\hat{\beta}^*)$$

$$\begin{aligned} \frac{d}{dk} MSE &= \frac{d}{dk} \Sigma \text{Var}(\hat{\beta}_i^*) + \frac{d}{dk} \text{Bias}^2(\hat{\beta}^*) \\ &= -2\sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \Sigma \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3} \\ &= -2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^3} (\sigma^2 - k\alpha_i^2) \end{aligned} \tag{14}$$

Based on Rolle's Theorem (Widder 1963), since

$$\lim_{k \to o} \frac{d}{dk} MSE = -2\sigma^2 \Sigma \frac{1}{\lambda_i^2} < 0$$

the optimal k is a positive value. Further, since

$$\frac{d^2}{dk^2} MSE = 6\sigma^2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^4} + 2 \Sigma \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^4} (\lambda_i - 2k) \tag{15}$$

and

$$\lim_{k \to o} \frac{d^2}{dk^2} MSE = 6\sigma^2 \Sigma \frac{1}{\lambda_i^3} + 2 \Sigma \frac{\alpha_i^2}{\lambda_i^2} > 0 \qquad (16)$$

And according to the theory of minimum (Widder, 1963), this positive value leads to the conclusion that a minimum exists for MSE (Kasarda and Shi, 1977).

7. Proof: from Eq. (14, footnote 6), at minimum MSE

$$\frac{d}{dk} MSE = -2 \Sigma \frac{\lambda_i}{(\lambda_i + k)^3} (\sigma^2 - k\sigma_i^2) = 0 \qquad (17)$$

Hence, the optimal k is a function of the true regression coefficient vector, α or β, and the variance of the residual of the linear regression model, i.e., σ^2.

8. Some of the more positive properties of ridge regression have been overlooked here. Note, however, that: (i) ridge regression produces smaller variances of the regression coefficients than OLS regression (for proof, see Lee 1979); (ii) ridge regression gives the minimum distance between $\hat{\beta}^*$ and the true vector β, which (in this sense) makes $\hat{\beta}^*$ a better estimator than $\hat{\beta}$, the OLS estimator (for proof, see Lee 1979); (iii) though ridge regression produces a smaller multiple R^2 than OLS regression (for proof, see Lee 1979), the ridge estimate gives a more accurate prediction equation than OLS regression if the bias introduced is not too large (for proof, see Lee 1979); and (iv) from (i) we can imply that ridge regression is less sensitive to sampling fluctuation.

9. These were: (1) the 5-factor Employee Satisfaction data from Price (1977); (2) the 8-factor Malmö data used by Bulcock et al. (1974) to investigate a problem in human capital theory; (3) the 10-factor example from Hoerl and Kennard (1970b); (4) the 10-factor mileage data from Hocking (1976); and (5) the 15-factor air pollution problem from McDonald and Schwing (1973).

10. On this point see Hauser, 1971. For important contributions to the school learning models literature see Carroll, 1963 and Bloom, 1976.

REFERENCES

ALLEN, D.M. (1974) "The relationship between variable selection and data augmentation and method for prediction." Technometrics 16: 125-127.

ALTHAUSER, R.P. (1971) "Multicollinearity and nonadditive regression models", pp. 453-472 in H.M. Blalock (ed.). Causal Models in the Social Sciences, Chicago: Aldine.

BIELBY, W.T. and R.M. HAUSER (1978) "Structural equation models", pp. 137-161 in A. Inkeles, J. Coleman, and N. Smelser (eds.) Annual Review of Sociology, Vol. 3, Palo Alto, CA: Annual Reviews, Inc.

BLALOCK, H.M. Jr. (1963) "Correlated independent variables: the problem of multicollinearity." Social Forces 42: 233-237.

BLALOCK, H.M. Jr. (1964) Causal Inferences in Non-experimental Research. Chapel Hill: University of North Carolina Press.

BLOOM, B.S. (1976) Human Characteristics and School Learning. New York: McGraw-Hill.

BULCOCK, J.W., I. FAGERLIND and I. EMANUELSSON (1974) Education and the socioeconomic career: U.S. - Swedish comparisons. Report No. 6, Institute for the Study of International Problems in Education, University of Stockholm.

BEEBE, M.J. and J.W. BULCOCK (1979) "Cueing strategies and basic skills in early reading", pp. 23-83 in J.W. Bulcock (ed.) Simple Ridge Regression with the Variance Normalization Criterion. Interim Report No. 1, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada, Ottawa, Ontario, Canada.

BULCOCK, J.W. and W.F. LEE (1979) "Cross-cultural differences in cognitive processes: differences in degree or of kind", p. 84-127 in J.W. Bulcock (Ed.) Simple Ridge Regression with the Variance Normalization Criterion. Interim Report No. 1, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada, Ottawa, Ontario, Canada.

BULCOCK, J.W., W.F. LEE, and W.S. LUK (1979) "A model of school learning using the variance normalization method of ridge regression analysis", pp. 156-202 in J.W. Bulcock (ed.) Simple Ridge Regression with the Variance Normalization Criterion. Interim Report No. 1, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada, Ottawa, Ontario, Canada.

CARROLL, J.B. (1963) "A model of school learning." Teachers College Record 64: 723-732.

CHRIST, C.F. (1966) Econometric Models and Methods. New York: John Wiley.

DEMPSTER, A.P., M. SCHATZOFF and N. WERMUTH (1977) "A simulation study of alternatives to ordinary least squares." Journal of the American Statistical Association 72: 77-91.

DUNCAN, O.D. (1975) "Path analysis: sociological examples." American Journal of Sociology 72: 3-16.

FARRAR, D.E. and R.R. GLAUBER (1967) "Multicollinearity in regression analysis: the problem revisited." Review of Economics and Statistics, 92-107.

FISHER, R.A. (1925) Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd.

FISHER, R.A. (1935) The Design of Experiments. Edinburgh: Oliver and Boyd.

GOLDBERGER, A. (1964) Econometric Theory. New York: John Wiley.

GORDON, R.A. (1968) "Issues in multiple regression." American Journal of Sociology 73: 592-616.

GUILKEY, D.K. and J.L. MURPHY (1975) "Directed ridge regression techniques in case of multicollinearity." Journal of the American Statistical Association 70: 769-775.

HAUSER, R.M. (1972) Socioeconomic Background and Educational Performance. Washington: American Sociological Association.

HEMMERLE, W.J. (1975) "An explicit solution for generalized ridge regression." Technometrics 17: 309-314.

HEMMERLE, W.J. and T.F. BRANTLE (1978) "Explicit and constrained generalized ridge estimation." Technometrics 20: 109-12[illegible].

HOCKING, R.R. (1976) "The analysis and selection of variables in linear regression." Biometrics 32: 1-49.

HOCKING, R.R., F.M. SPEED and M.J. LYNN (1976) "A class of biased estimators in linear regression." Technometrics 12: 55-67.

HOERL, A.E. and R.W. KENNARD (1970a) "Ridge regression: biased estimation for nonorthogonal problems." Technometrics 12: 55-67.

HOERL, A.E. and R.W. KENNARD (1970b) "Ridge regression: application to nonorthogonal problems." Technometrics 12: 69-82.

HOERL, A.E. and R.W. KENNARD (1976) "Ridge regression: iterative estimation of the biasing parameter." Communications in Statistics - Theoretical Methods A5(1): 77-88.

HOERL, A.E., R.W. KENNARD, and K.F. BALDWIN (1975) "Ridge regression: some simulation." Communications in Statistics, 4(2): 105-123.

HUI, B.S. (1978) Soft Modelling by the Partial Least Squares Approach. Ph.D. dissertation, University of Pennsylvania (unpublished).

JOHNSTON, J. (1963) Econometric Methods. New York: McGraw-Hill.

JOHNSTON, J. (1972) Econometric Methods. Second Edition. New York: McGraw-Hill.

JORESKOG, K.G. (1970) "A general method for analysis of covariance structures." Biometrica 57: 239-51.

JORESKOG, K.G. (1973) "A general method for estimating a linear structural equation system", pp. 85-112 in A.S. Goldberger and O.D. Duncan (eds.) Structural Equation Models in the Social Sciences. New York: Seminar Press.

JORESKOG, K.G. (1976) "Structural equation models in the social sciences: specification, estimation, and testing." Symposium in Applied Statistics, Dayton, Ohio.

JORESKOG, K.G. (1978) "Structural analysis of covariance and correlation matrices", Psychometrica 43: 443-477.

KASARDA, J.D. and W.F. SHIH (1977) "Optimal bias in ridge regression approaches to multicollinearity." Sociological Methods and Research 5: 461-470.

KING, D. Feedback and Reciprocal Causation: A Monte Carlo Study of the Relative Small Sample Properties of Small Sample Estimators. Ph.D. dissertation, University of North Carolina (unpublished).

KISH, L. (1975) "Representation, randomization, and control", pp. 261-284 in H.M. Blalock et al. Mathematic Models and International Perspectives on Mathematical Sociology. New York: Academic Press.

KMENTA, J. (1971) Elements of Econometrics. New York: MacMillan.

LAND, K.C. and M. FELSON (1978) "Sensitivity analysis of arbitrarily identified simultaneous-equation models." Sociological Methods and Research 6: 283-307.

LAWLESS, J.F. and P. WANG (1976) "A simulation study of ridge and other regression estimators." Communications in Statistics-Theoretical Methods A5(4): 307-323.

LEE, W.F. (1979) Model Estimation Using Ridge Regression with the Variance Normalization Criterion. Interim Report No. 2, Grant No. S76-0645. The Social Sciences and Humanities Research Council of Canada, Ottawa, Ontario, Canada.

LEE, W.F. and W.S. LUK (1979) "Simple ridge regression with the variance normalization criterion", pp. 128-155 in J.W. Bulcock (ed.), Simple Ridge Regression with the Variance Normalization Criterion. Interim Report No. 1, Grant No. S76-0645, The Social Sciences and Humanities Research Council of Canada, Ottawa, Ontario, Canada.

LINDLEY, D.V. and A.F.M. SMITH (1972) "Bayes estimates for the linear model." Journal of the Royal Statistical Society, Series B, 34: 1-41.

MARQUARDT, D.W. (1970) "Generalized inverse, ridge regression, biased linear estimation, and nonlinear estimation." Technometrics 12: 591-612.

MARQUARDT, D.W. and R.D. SNEE (1975) "Ridge regression in practice." The American Statistician 29: 3-20.

MCDONALD, G.C. and D.I. GALARNEAU (1975) "A Monte Carlo evaluation of some ridge-type estimators." Journal of the American Statistical Association 70: 407-416.

MCDONALD, G.C. and R.C. SCHWING (1973) "Instabilities of regression estimates relating air pollution to mortality." Technometrics 15: 463-481.

NAMBOODIRI, N.K., L.F. CARTER, and H.M. BLALOCK jr. (1975) Applied Multivariate Analysis and Design. New York: McGraw-Hill.

NOONAN, R. and H. WOLD (1977) "NIPALS path modelling with latent variables: analyzing school survey data using nonlinear, iterative, partial least squares." Scandinavian Journal of Educational Research 21: 33-61.

NOONAN, R. and H. WOLD (1978) "PLS path modelling with latent variables: analyzing school survey data using partial least squares - part II." Working paper no. 1, Institute of International Education, University of Stockholm, Sweden.

OBENCHAIN, R.L. (1975) "Ridge analysis following a preliminary test of the shrunken hypothesis." Technometrics 17: 431-445.

PINDYCK, R.S. and D.L. RUBINFELD (1976) Econometric Methods and Economic Forecasts. New York: McGraw-Hill.

PRICE, B. (1977) "Ridge regression: application to non-experimental data." Psychological Bulletin 84: 759-766.

SUMMERS, R. (1965) "A capital intensive approach to the small sample properties of various simultaneous equation estimators." Econometrics 33: 1-41.

THEIL, H. (1971) Principles of Econometrics. New York: Wiley.

TINBERGEN, J. (1939) Statistical Testing of Business Cycle Theories II: Business in the United States of America 1919-1932. Geneva: The League of Nations.

VINOD, H.D. (1976) "Application of new ridge regression methods to a study of Bell system scale economics. Journal of the American Statistical Association 71: 835-841.

WANNACOTT, R.J. and J.H. WANNACOTT (1969) Econometrics. New York: Wiley.

WHITE, K.J. (1978) SHAZAM Version 2.1. Department of Economics, Rice University, Houston, Texas, U.S.A.

WIDDER, D.V. (1963) Advanced Calculus. Englewood Cliffs, New Jersey: Prentice-Hall.

WOLD, H. (1975) "Path models with one or two latent variate aggregates", in H.M. Blalock Jr. et al. Quantitative Sociology. New York: Seminar Press.

WOLD, H. (1977a) "On the transition from pattern recognition to model building", pp. 536-549 in R. Henn and O. Moeschlin (eds.), Mathematical Economics and Game Theory: Essays in Honor of Oskar Morgenstern. Berlin: Springer.

WOLD, H. (1977b) "Open path models with latent variables", in H.M. Schoenfeld and A. Moeschlieu (eds.), Essays in Honour of Wilhelm Krelle. Tubingen: Mohr.

WOLD, H. (1977c) "Soft modelling: intermediate between traditional model building and data analysis", in Banach Publications 6: Mathematical Statistics. Warsaw, Poland.

WOLD, H. (1978) Ways and Means of Multidisciplinary Studies. Report No. 3, Department of Statistics, University of Uppsala, Uppsala, Sweden.